


Computer-Aided Diagnosis of Prostate Cancer Using a Deep Convolutional Neural Network From Multiparametric MRI

Yang Song, BS ,¹ Yu-Dong Zhang, MD, PhD,^{2*} Xu Yan, PhD,³ Hui Liu, PhD,⁴ Minxiong Zhou, PhD,⁵ Bingwen Hu, PhD,¹ and Guang Yang, PhD^{1*}

Background: Deep learning is the most promising methodology for automatic computer-aided diagnosis of prostate cancer (PCa) with multiparametric MRI (mp-MRI).

Purpose: To develop an automatic approach based on deep convolutional neural network (DCNN) to classify PCa and noncancerous tissues (NC) with mp-MRI.

Study Type: Retrospective.

Subjects: In all, 195 patients with localized PCa were collected from a PROSTATEx database. In total, 159/17/19 patients with 444/48/55 observations (215/23/23 PCas and 229/25/32 NCs) were randomly selected for training/validation/testing, respectively.

Sequence: T₂-weighted, diffusion-weighted, and apparent diffusion coefficient images.

Assessment: A radiologist manually labeled the regions of interest of PCas and NCs and estimated the Prostate Imaging Reporting and Data System (PI-RADS) scores for each region. Inspired by VGG-Net, we designed a patch-based DCNN model to distinguish between PCa and NCs based on a combination of mp-MRI data. Additionally, an enhanced prediction method was used to improve the prediction accuracy. The performance of DCNN prediction was tested using a receiver operating characteristic (ROC) curve, and the area under the ROC curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated. Moreover, the predicted result was compared with the PI-RADS score to evaluate its clinical value using decision curve analysis.

Statistical Test: Two-sided Wilcoxon signed-rank test with statistical significance set at 0.05.

Results: The DCNN produced excellent diagnostic performance in distinguishing between PCa and NC for testing datasets with an AUC of 0.944 (95% confidence interval: 0.876–0.994), sensitivity of 87.0%, specificity of 90.6%, PPV of 87.0%, and NPV of 90.6%. The decision curve analysis revealed that the joint model of PI-RADS and DCNN provided additional net benefits compared with the DCNN model and the PI-RADS scheme.

Data Conclusion: The proposed DCNN-based model with enhanced prediction yielded high performance in statistical analysis, suggesting that DCNN could be used in computer-aided diagnosis (CAD) for PCa classification.

Level of Evidence: 3

Technical Efficacy: Stage 2

J. MAGN. RESON. IMAGING 2018;48:1570–1577.

Magnetic resonance imaging (MRI) is gradually evolving into an important tool for prostate cancer (PCa).^{1,2} Studies have generally focused on the use of multiparametric MRI (mp-MRI) sequences, such as T₂-weighted imaging

(T₂WI), diffusion-weighted imaging (DWI), and dynamic contrast-enhanced (DCE) imaging for lesion detection, differentiation, and staging. Although MRI images feature high spatial resolution, the signal intensity on MRI images might

View this article online at wileyonlinelibrary.com. DOI: 10.1002/jmri.26047

Received Jan 9, 2018, Accepted for publication Mar 23, 2018.

*Address reprint requests to: Y.-D.Z., Department of Radiology, First Affiliated Hospital with Nanjing Medical University, 300 Guangzhou Road, Nanjing, China. E-mail: njmu_zyd@163.com or G.Y., Shanghai Key Laboratory of Magnetic Resonance, East China Normal University, 3663 N. Zhongshan Road, Shanghai, China, 200062. E-mail: gyang@phy.ecnu.edu.cn

From the ¹Shanghai Key Laboratory of Magnetic Resonance, East China Normal University, Shanghai, China; ²Department of Radiology, First Affiliated Hospital with Nanjing Medical University, Nanjing, China; ³MR Scientific Marketing, Siemens Healthcare, Shanghai, China; ⁴Research Department, HImagingTek Ltd., Shanghai, China; and ⁵Shanghai University of Medicine & Health Sciences, Shanghai, China

exhibit overlap among cancer, postbiopsy hematoma, benign prostatic hypertrophy, fibrosis, and prostatitis. Conventional methods, ie, semiquantitative, quantitative, and structured report schemas, have inconsistent reproducibility. The Prostate Imaging Reporting and Data System (PI-RADS) was designed to standardize mp-MRI sequence acquisition and reporting and increase the sensitivity of PCa identification by introducing a 5-score strategy to reveal the likelihood of clinically significant disease: highly unlikely (score 1), unlikely (score 2), equivocal (score 3), likely (score 4), and highly likely (score 5).³ The accuracy of PI-RADS largely depends on the experience of the radiologist, and interobserver variation also exists. Therefore, advanced methods incorporating mp-MRI and quantitative analysis that favor reproducibility and accuracy are of high priority.

Over the past two decades, computer-aided diagnosis (CAD) has been a rapidly growing and active area of research. Several CAD systems for the radiological assessment of various malignancies in the breast,⁴ lung,⁵ and prostate⁶ have been developed, and encouraging results have been reported. Studies incorporating large numbers of imaging features have demonstrated the feasibility of CAD for aiding the discrimination of benign from malignant lesions in the prostate with high accuracy. In most feature-based approaches, the conventional CAD system consists of a series of processing steps, such as lesion identification, segmentation, feature selection, and model-dependent classification, leading to a variety of findings due to methodological inconformity.

Deep convolutional neural networks (DCNNs) are a newly emerging form of CAD analysis that allow the automatic extraction of features and supervised learning of large amounts of data to form quantitative decisions. Growing evidence indicates that deep learning analysis might be a potential alternative to conventional hand-crafted methodologies for solving pattern recognition and imaging classification problems.^{7–10} However, studies focused on combining DCNNs and mp-MRI to solve lesion detection and classification problems of the prostate have been limited.^{11,12} Le et al explored three traditional DCNN models (VGGNet, GoogleNet, and ResNet) using T₂WI, apparent diffusion coefficient (ADC), and their combinations as the inputs. The augmentation method was used to increase the number of samples in the training dataset.¹¹ Liu et al proposed XmasNet, which was inspired by VGGNet. Batch normalization was added after convolution to accelerate the convergence during back-propagation.¹² In these studies, patches (a square region) were extracted from images as the input of the DCNN model to help the DCNN focus on the region of interest (ROI) of the mp-MR images and to increase the number of samples in the dataset.

This study aimed to assess the value of a DCNN model-based CAD system in diagnosing PCa using mp-

MRI. We proposed an optimized patch-based DCNN model based on VGGNet and integrated an enhanced prediction method into the prediction stage to increase the prediction accuracy. Additionally, we performed a head-to-head comparison of our DCNN scheme with the use of PI-RADS by an experienced radiologist to differentiate benign from malignant prostate lesions, and also evaluated the performance of their combination.

Materials and Methods

Subjects

Mp-MRI data were downloaded from the Cancer Imaging Archive (TCIA) data portal for SPIE-AAPM-NCI PROSTATEx challenge (<https://doi.org/10.7937/K9TCIA.2017.MURS5CL>). The PROSTATEx portal included data from 204 patients with 330 PCa findings. Test data from another 140 patients were not used in our study. For each of the PCa findings, the image location in the MRI data was provided. In this study, data from 195 patients with 261 PCa regions with good image quality and consistent imaging parameters were carefully selected from the total of 204 datasets and used in the analysis. For DCNN training and testing, 286 noncancer (NC) regions were labeled by a dedicated radiologist (see details in the DCNN Training subsection).

We randomly selected 159 patients with 444 ROIs (215 PCas and 229 NCs) for DCNN mode training, 17 patients with 48 ROIs (23 PCas and 25 NCs) for validation, and the remaining 19 patients with 55 ROIs (23 PCas and 32 NCs) for testing. The validation data were monitored to evaluate the weights to help update the learning rate and allow early stopping of training to avoid overfitting. A flow chart of the study design is shown in Fig. 1.

Acquisition Protocol

The axial T₂WI, DWI (b-value of 50, 400, and 800 s/mm²) and ADC maps were used for imaging analysis. All data were acquired on 3T MR scanners with pelvic phased-array coils (Magnetom Trio and Skyra, Siemens Healthcare, Erlangen, Germany). Axial T₂WIs were acquired using a turbo-spin-echo sequence with an in-plane resolution of $\sim 0.5 \times 0.5$ mm² and a slice thickness of 3.6 mm. The DWI series was acquired with a single-shot echo-planar-imaging sequence, an in-plane resolution of 2 mm, and a slice thickness of 3.6 mm.¹³ DCE and K^{trans} images were excluded given their limited attributions for cancer diagnosis based on PI-RADS v2 guidelines.³

DCNN Model

We referred to VGGNet with a 1×1 layer and designed our patch-based DCNN architecture (Fig. 3).¹⁴ Our model contained three blocks followed by three fully connected layers. Each block included two convolutional layers with a filter size of 3×3 , one convolutional layer with a filter size of 1×1 , and a max-pooling layer with a filter size of 2×2 . Compared with VGGNet, the layer number was reduced since the size of the patch is smaller and the task is relatively simple. Dropout was performed before max pooling and a fully connected layer to avoid overfitting. Batch normalization was used after each convolutional process to accelerate

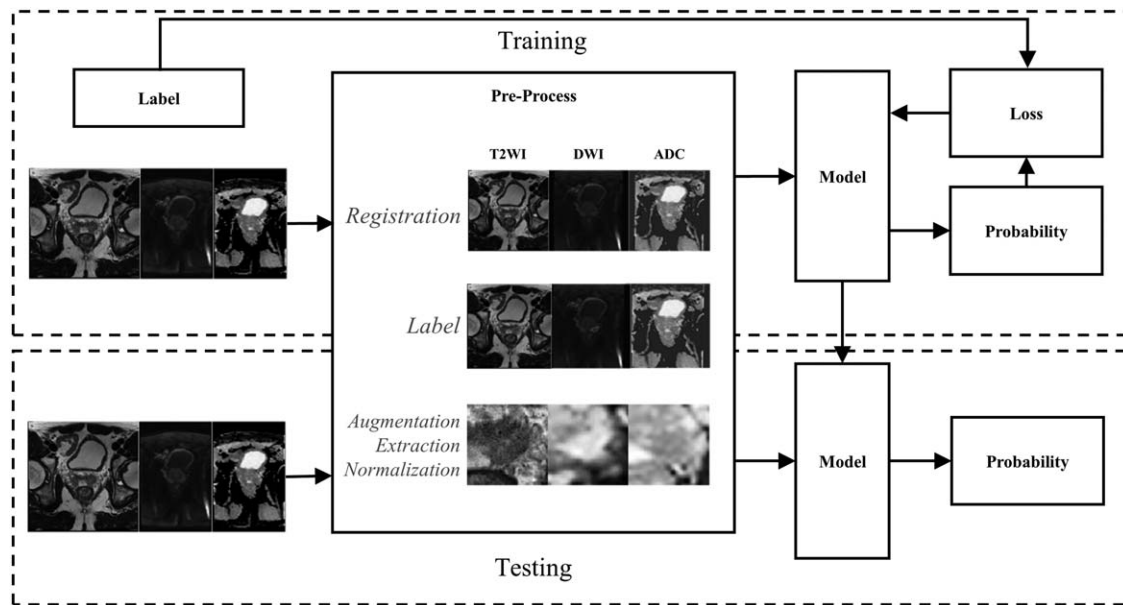


FIGURE 1: The flow-chart of the process of the model including training step and testing step.

the convergence during back-propagation.¹⁵ We used an exponential linear unit instead of a rectified linear unit to update the weights even when the activation was negative. We added a 1×1 layer with two filters followed by soft-max activation at the end of the net to obtain the probability map of PCa and NC. We used the ADAM algorithm to update the weights with an initial learning rate of 0.001 and cross-entropy as our cost function.¹⁶ The architecture of the model was built by Keras (v. 2.0.8) and TensorFlow (v. 1.3.0), and the model was trained on a NVIDIA TITAN Xp graphic card.

DCNN Training

The data were preprocessed using the following steps before import into the model (Fig. 1).

REGISTRATION. Low-resolution DWI/ADC images were aligned and resampled to high-resolution T₂W images using an established registration toolbox for rigid and nonrigid transformation (Elastix).¹⁷ The parameter configuration for the registration was set according to Klein et al's work.¹⁷

ROI LABELING. Since PROSTATEx only provides the position of biopsy with cancerous results, locating the center position of the PCa region is necessary for subsequent image analysis. Therefore, a dedicated radiologist (Y.Z.) with 11 years of experience in interpreting cross-sectional genitourinary images was recruited to relabel the entire contours of the cancer lesions with reference to the provided cancer position. Moreover, two representative ROIs were randomly drawn on normal peripheral zone (PZ) and transition zone (TZ) regions as a control (NC regions) by avoiding the PCa region reported in PROSTATEx. The ROI was drawn semiautomatically using dedicated in-house software written in Python 3.6.1 (<https://www.python.org>). A rough ROI was first manually contoured by the radiologist to cover the entire lesion. Then the lesion and surrounding normal tissue were automatically separated using an

iterative threshold method.¹⁸ Finally, the dedicated radiologist manually verified and fine-tuned the ROI border slice-by-slice (Fig. 2).

AUGMENTATION AND PATCH EXTRACTION. Given an ROI-labeled image, we augmented the images using the following strategies: random rotation from -20 to 20° , random shift of less than 2 pixels along the x-axis and y-axis, random stretching from 0.9 to 1.1 times along the x-axis and y-axis, and random horizontal flipping of the image. Then a 2D patch with a size of 65×65 centered at the ROI center was extracted from the transverse plane from all sequences.

NORMALIZATION. All sequences were concatenated as separate channels to be input into the DCNN net. To reduce variations in signal intensity among different images, all data were normalized before training. The general normalization of digital images in DCNN involves subtracting the mean value of all images and then dividing the images by the standard deviation. In our case, normalization was applied separately to each sequence image of each patient because the contrast difference between images of different sequences is intrinsic and contains useful information for diagnosis. Thus, the differences in image contrast should not be removed by normalization.

Enhanced Prediction

To improve the accuracy and robustness of prediction, we incorporated a so-called enhanced prediction approach with the trained DCNN model. This approach is based on the same idea as the augmentation process in the training step. For each test sample, we augmented the data 11 times (random rotation, shift, stretch, and horizontal flip) and predicted each test sample separately by the same model. The average probability was used in the final prediction. The enhanced prediction was applied in the prediction of the testing data.

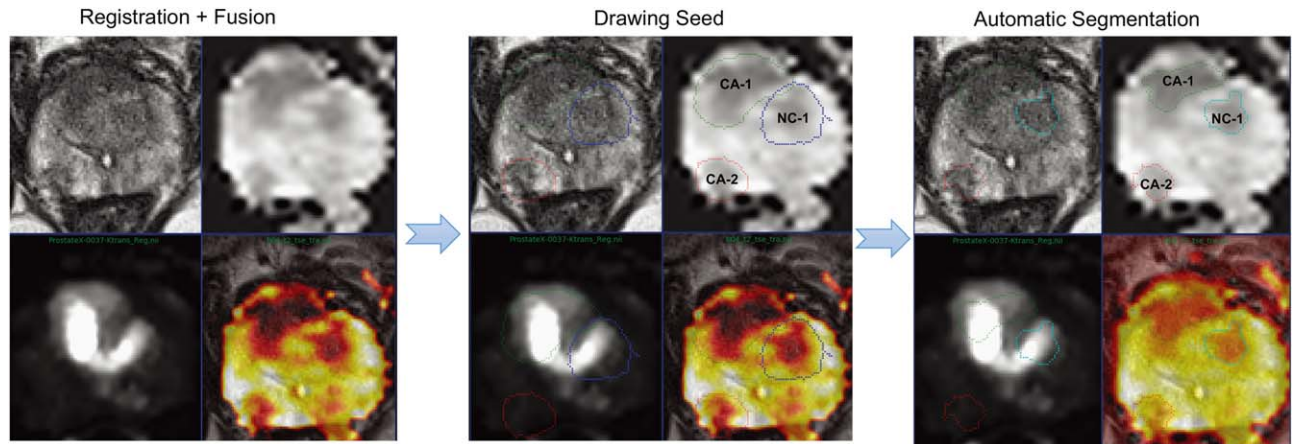


FIGURE 2: Representative case to show imaging registration, fusion, placement of ROI. Image registration and fusion was first performed with an in-house software package (left). Then the seeds of ROIs were drawn generally on the bases of location information provided by TCIA data portal (middle), in the present case; CA-1 and CA-2 represent the cancer regions. Additionally, a suspicious regional seed (NC-1) was set at transition zone as a control. Lastly, the outer margin of each ROI was automatically segmented and fine-tuned by the dedicated radiologist (right).

DCNN Evaluation

To evaluate the performance of the enhanced prediction approach, we compared this approach with the traditional prediction approach using the test data. The same DCNN model acquired in the training step was used for both approaches. The combination of T₂-ADC-DWI (b-value = 800 s/mm²) images was adopted as the model input.

To explore the performance of the model input with different sequence combinations, we trained three DCNN models independently with three inputs: including T₂-ADC images, T₂-ADC-DWI (only b-value = 800 s/mm²) images, and T₂-ADC-DWI (three different b-values) images. The results of these three models were compared.

We also explored the influence of the block number setting on our DCNN model. We fixed the input with T₂-ADC-DWI (b-value = 800 s/mm²) images and created three DCNN models with two, three, or four blocks. All models followed the same fully connected layers.

Additionally, we evaluated the clinical usefulness of adding the value from our DCNN model to PI-RADS. The diagnostic performances of the PI-RADS scheme, DCNN model, and a joint

model using PI-RADS and DCNN were compared. All lesions were reinterpreted by the dedicated radiologist, who used a 5-point PI-RADS scoring scheme.

Statistical Analysis

Validation of the prediction models comprised two activities. First, the performance of PCa and NC discrimination of the model was quantified using receiver operating characteristic (ROC) curve analysis. Paired comparisons of the area under the ROC curves (AUC) were performed using the two-sided Wilcoxon signed-rank test. The sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated at a cutoff value that maximized the value of the Youden index. Second, the cost-effect benefit analysis proposed by Vickers and Elkin was performed to evaluate the additional benefits of the DCNN method when combined with the conventional PI-RADS scheme.¹⁹ Statistical analysis was conducted using the R package (Risk Model Decision Analysis, <https://github.com/mdbrown/rmda>) and a python module (scikit-learn, <http://scikit-learn.org>). The reported statistical significance levels were all two-sided, and statistical significance was set at 0.05.

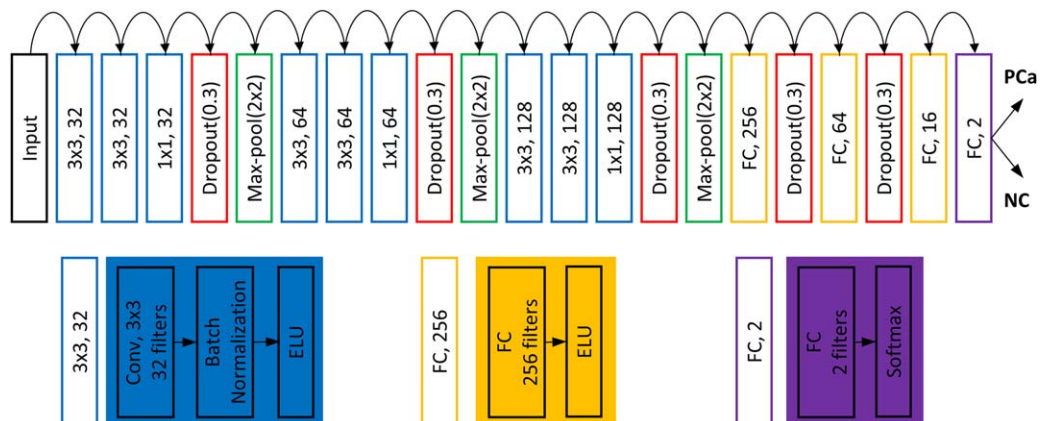


FIGURE 3: The architecture of the proposed network.

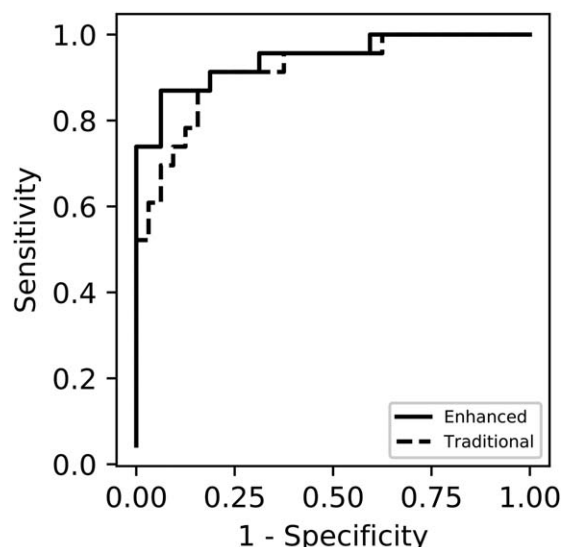


FIGURE 4: ROC comparison of two prediction models for discriminating CA from NC tissue. Solid and dashed line represents the enhanced prediction and traditional prediction, respectively.

Results

Traditional Prediction vs. Enhanced Prediction

The traditional and enhanced prediction results of our DCNN model were compared based on their discrimination ability (Fig. 4). The ROC analysis showed that the AUC of the enhanced prediction model (0.944, 95% confidence interval [CI]: 0.876–0.994) was significantly higher than that of the traditional prediction model (0.917, 95% CI: 0.841–0.978; $P < 0.001$).

The detailed diagnostic performance of the two prediction approaches for the testing dataset is summarized in Table 1. Enhanced/traditional prediction was evaluated in terms of sensitivity (87.0%/91.3%), specificity (90.6%/81.2%), PPV (87.0%/77.8%), and NPV (90.6%/92.9%), with optimal cutoff values of 0.92/0.85 that maximized the Youden index.

Optimized Combination of Imaging Sequences

The ROCs of the DCNN models with different input combinations are summarized in Fig. 5. The model combining

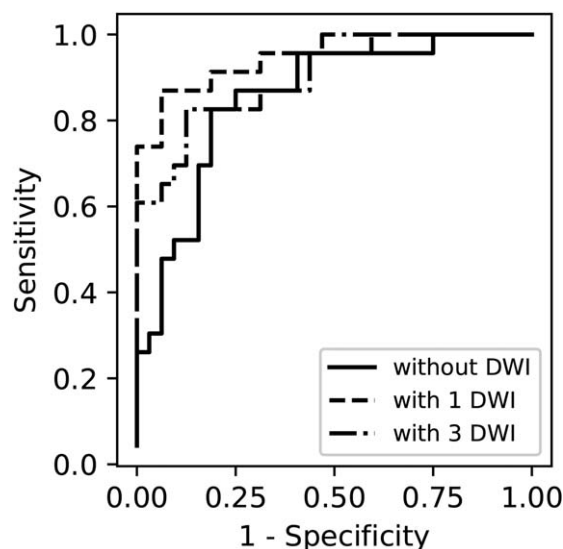


FIGURE 5: The ROC curve of the model with different imported data. Solid, dashed, and dash-dot lines denote the ROC curve of the model with imported data including T₂W-ADC, 2W-ADC-DWI (high b value), and T₂W-ADC-DWI (three different b values), respectively.

T₂W, ADC, and high b-value DWI (AUC: 0.944, 95% CI: 0.876–0.994) outperformed the model combining T₂W, ADC, and all b-value DWI (AUC: 0.905, 95% CI: 0.816–0.971) and the model combining T₂W and ADC (AUC: 0.853, 95% CI: 0.750–0.943), with significantly higher discrimination accuracy ($P < 0.001$).

Optimized Block Number

The influence of blocks on model performance is summarized in Fig. 6. The model using three blocks resulted in the best performance (AUC: 0.944, 95% CI: 0.876–0.994) compared with the model with two blocks (AUC: 0.909, 95% CI: 0.819–0.977) and model with four blocks (AUC:

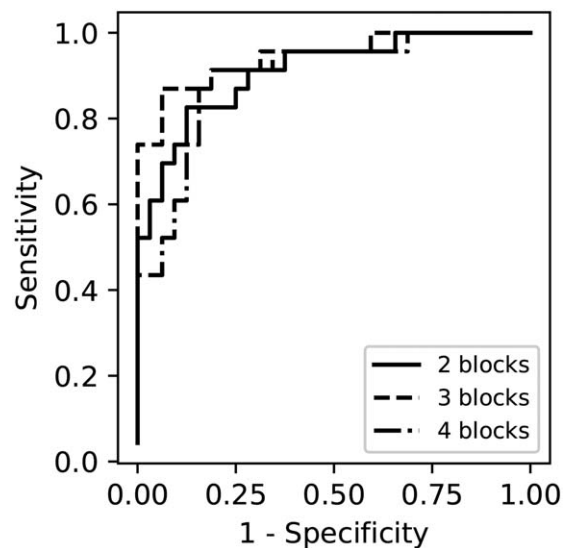


FIGURE 6: ROC curve of different model. Solid, dashed, and dash-dot lines denote the model with two, three, and four blocks, respectively.

TABLE 1. Diagnosis Details for the Same Model With or Without Enhancement of the Prediction Process

		Traditional prediction		Enhanced prediction	
		CA	NC	CA	NC
Biopsy	CA	21	2	20	3
Result	NC	6	26	3	29

The threshold value was chosen according to the Youden index. The threshold value was set to 0.85 for traditional prediction and 0.92 for enhanced prediction.

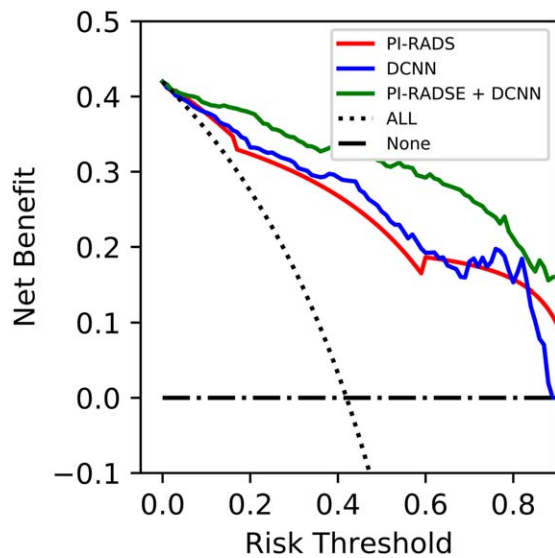


FIGURE 7: Decision curve analysis of PI-RADS scheme, DCNN model, and PI-RADS + DCNN. The y-axis measures the net benefits, wherein the red, blue, and green line represents PI-RADS, DCNN, and PI-RADS + DCNN, respectively. The dotted line represents the assumption that all patients have CA. The dash-dot line represents the assumption that no patient has CA. It shows that, if the risk threshold probability is set between 5% and 80%, adding DCNN to the PI-RADS scheme results in more benefit for diagnostic performance.

0.897, 95% CI: 0.805–0.965). Based on the optimal cutoff values that maximized the Youden index, the model with three blocks was superior to any other models with respect to discrimination accuracy ($P < 0.001$).

Added Value of DCNNs

Decision curve analysis of the conventional expert-based PI-RADS model, DCNN model, and the joint model using PI-RADS scores and DCNN outputs was performed (Fig. 7). The estimated net benefit of the DCNN model was similar to that of the PI-RADS scheme. By adding DCNN to the PI-RADS scheme, increased diagnostic performance was obtained if the risk threshold probability was set between 5% and 80%.

Discussion

In this study we developed a novel DCNN model with mp-MRI and evaluated its performance for differentiating benign and malignant prostate lesions based on the PROSTATEx database. An enhanced prediction method, optimized combination of imaging sequences, and optimized block number of the DCNN model were additionally evaluated. The results showed that our DCNN model with the enhanced prediction method, combination of T_2W , ADC, and high b-value DWI data and three-block architecture achieved the highest diagnostic accuracy in terms of sensitivity, specificity, PPV, and NPV for discriminating PCa from benign lesions. Additionally, we performed a head-to-head comparison between PI-RADS and DCNNs. The

combination of DCNN with the PI-RADS scheme provided notable gains compared with the PI-RADS and DCNN model for cancer classification and thus improved cancer diagnosis and the decision-making process.

Several studies used DCNN on the prostate PCa with mp-MR images. Le et al introduced data augmentation into data preparation to increase the performance of the DCNN model.¹¹ This data augmentation strategy was also used in our study. Furthermore, we augmented each sample in the testing dataset and used majority voting for the classification. Liu et al trained several DCNN models and each took one combination of different sequence images as input and a weighted majority voting was used for classification.¹² Interestingly, T_2W -ADC-DWI was assigned the largest weight among all combinations, which was consistent with our results. Classical DCNN models proved to work on the classification of the visual images. They could be used on a medical image directly.¹¹ Our work focused on the classification of prostate PCa and NC with mp-MR images. An adaptive architecture of the DCNN model could be designed based on the features of mp-MR images and prostate PCas.

The enhanced prediction method in our DCNN network was inspired by data augmentation, which is commonly used in data preprocessing for deep learning to increase the robustness of the model, especially when the size of training data is limited.²⁰ We demonstrated that this data augmentation strategy could also be used in the prediction stage to significantly improve the prediction results, and we therefore called this strategy “enhanced prediction.” This concept is similar to the diagnosis of the same data several times by a single doctor, each time with a different view angle. This is different from the ensemble learning, which builds several models to obtain better predict performance.²¹ Enhanced prediction can save the time of building models and improve the predictive accuracy. The learned DCNN model predicted these images separately and determined the final diagnosis. Enhanced prediction can reduce the likelihood of an improper diagnosis.

For the combinations of different imaging sequences, the models involving DWI exhibited higher specificity. The features of the cancerous region with a high signal in high b-value DWI were learned easily by the DCNN than the models including only T_2 and ADC images,¹² consistent with the PI-RADS v2 guidelines.³ The model using one high b-value DWI image achieved a better result than the model using three DWI images of different b-values, which may be due to the architecture of the model. In the model, mp-MRI images and their corresponding feature maps were combined and merged from the first layer, making it difficult to separate these feature maps in the following convolution layers. Thus, a low b-value and feature maps that contain less lesion information may act as noise terms in the model, reducing the performance of the DCNN model.

Numerous hyperparameters in the DCNN model influence the prediction outcome, such as the depth of the network. The number of hyperparameters was set according to the objective problem and the number of samples for training. According to DCNN theory, the deeper the DCNN, the larger receptive field of it and the greater its ability to extract abstract information from images.²² A deeper DCNN requires more samples from the training dataset to learn more hyperparameters. The model with three blocks yielded more accurate predictions than the model with two or four blocks. The model with three blocks may represent a good balance between sufficient abstract information (more than the two-block model) and a suitable sample size for hyperparameter training (smaller parameter size than the four-block model). From another perspective, compared with the model with three blocks, the models with two and four blocks resulted in underfitting and overfitting, respectively, in diagnosing PCa from NC.²³

Some limitations of our study warrant mention. 1) The number of patients used for model training was less than 200. More data would yield better results. The size of our dataset might be suboptimal for deep learning. 2) The training process is sensitive to the parameter initialization. Compared with random parameter initialization, proper parameter initialization together with a fine-tuning strategy may help better train the model in less time.²⁴ 3) The model was trained on data with good image quality and consistent imaging parameters. For clinical application, a good model should be sufficiently robust to accommodate mp-MRI with slightly different imaging parameters. More dataset samples with different imaging parameters could be used to train a more widely applicable model. Generative adversarial networks could be another choice to synthesize mp-MR images to increase the number of dataset samples.²⁵

In conclusion, we have proposed a patch-based DCNN model to classify prostate PCa and NC from mp-MR images. Enhanced prediction was integrated into the prediction stage to improve the performance of the DCNN model. The high performance of our DCNN model in statistical analysis supports its potential to improve PCa diagnosis in the CAD system with high sensitivity to propose suspected regions for further diagnosis.

Acknowledgment

Contract grant sponsor: Key Project of the National Natural Science Foundation of China; contract grant number: 61731009 (to G.Y.); Contract grant sponsor: China Postdoctoral Fund; contract grant number: 2015M580453 (to Y.D.Z.); Contract grant sponsor: Key Social Development Program for the Ministry of Science and Technology of Jiangsu Province; contract grant number: BE2017756 (to Y.D.Z.)

Conflict of Interest

Xu Yan, from a commercial company, Siemens Healthcare, was an MR collaboration scientist doing technical support in this study under Siemens collaboration regulation without any payment and personal concern regarding this study. Hui Liu, a collaboration scientist from hImagingTek Ltd, gave technical support in this study without any payment and personal concern regarding.

References

- de Rooij M, Hamoen EHJ, Fütterer JJ, Barentsz JO, Rovers MM. Accuracy of multiparametric mri for prostate cancer detection: a meta-analysis. *Am J Roentgenol* 2014;202:343–351.
- Langer DL, van der Kwast TH, Evans AJ, Trachtenberg J, Wilson BC, Haider MA. Prostate cancer detection with multi-parametric MRI: logistic regression analysis of quantitative T2, diffusion-weighted imaging, and dynamic contrast-enhanced MRI. *J Magn Reson Imaging* 2009;30:327–34.
- Weinreb JC, Barentsz JO, Choyke PL, et al. PI-RADS Prostate Imaging — Reporting and Data System: 2015, Version 2. *Eur Urol* 2016;69:16–40.
- Dromain C, Boyer B, Ferré R, Canale S, Delaloge S, Balleyguier C. Computed-aided diagnosis (CAD) in the detection of breast cancer. *Eur J Radiol* 2013;82:417–423.
- Awai K, Murao K, Ozawa A, et al. Pulmonary nodules at chest CT: effect of computer-aided diagnosis on radiologists' detection performance. *Radiology* 2004;230:347–352.
- Lemaître G, Martí R, Freixenet J, Vilanova JC, Walker PM, Meriaudeau F. Computer-aided detection and diagnosis for prostate cancer based on mono and multi-parametric MRI: A review. *Comput Biol Med* 2015;60:8–31.
- Schmidhuber J. Deep learning in neural networks: An overview. *Neural Networks* 2015;61:85–117.
- Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. *Annu Rev Biomed Eng* 2017;19:221–248.
- Alipanahi B, Delong A, Weirauch MT, Frey BJ. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat Biotechnol* 2015;33:831–838.
- Aerts HJWL, Velazquez ER, Leijenaar RTH, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* 2014;5:4006.
- Le MH, Chen J, Wang L, et al. Automated diagnosis of prostate cancer in multi-parametric MRI based on multimodal convolutional neural networks. *Phys Med Biol* 2017;62:6497–6514.
- Liu S, Zheng H, Feng Y, Li W. Prostate cancer diagnosis using deep learning with 3D multiparametric MRI. Volume 10134. In Armato SG, Petrick NA, editor. *International Society for Optics and Photonics*; 2017:1013428.
- Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H. Computer-aided detection of prostate cancer in MRI. *IEEE Trans Med Imaging* 2014;33:1083–1092.
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. eprint arXiv:1409.1556, 2014.
- Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. eprint arXiv:1502.03167, 2015.
- Kingma DP, Ba J. Adam: A method for stochastic optimization. eprint arXiv:1412.6980, 2014.
- Klein S, Staring M, Murphy K, Viergever MA, Pluim JPW. elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging* 2010;29:196–205.

18. Gonzalez RC, Woods RE. Digital image processing, 2nd ed. Englewood Cliffs, NJ: Prentice Hall; 2008.
19. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;26:565–574.
20. Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. eprint arXiv:1511.06434, 2015.
21. Kuncheva LI, Whitaker CJ. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Mach Learn* 2003;51:181–207.
22. Coates A, Ng AY. Selecting receptive fields in deep networks. *The Neural Information Processing Systems Conference*, 2011;2528–2536.
23. Michalski RS, Carbonell JG, Mitchell TM. *Machine learning?: An artificial intelligence approach*. Berlin, Heidelberg: Springer; 1983.
24. Rusu AA, Rabinowitz NC, Desjardins G, et al. Progressive neural networks. eprint arXiv:1606.04671, 2016.
25. Dar SUH, Yurt M, Karacan L, Erdem A, Erdem E, Çukur T. Image synthesis in multi-contrast MRI with conditional generative adversarial networks. eprint arXiv:1802.01221, 2018.